PySpark Groupby agg() is used to calculate more than one aggregate at a time on grouped DataFrames.

from pyspark.sql import \*  
from pyspark.sql.functions import \*  
  
spark = SparkSession.builder.appName('groupBy\_agg').getOrCreate()  
  
myData = [(1,'Harsha','male',2000,'IT'),  
 (2,'Mokshit','male',3000,'HR'),  
 (3, 'Harika', 'female', 3500, 'payroll'),  
 (4, 'Kiran', 'male', 4000, 'HR'),  
 (5, 'venkat', 'male', 5000, 'IT'),  
 (6, 'Rama', 'female', 4500, 'payroll')  
 ]  
  
mySchema = ['id','name','gender','salary','dept']  
  
df = spark.createDataFrame(myData,mySchema)  
  
df.groupBy(df.dept).agg(count('\*').alias('countOfEmp'),  
 min('salary').alias('min\_sal'),  
 max('salary').alias('max\_sal')).show()

**Output:**

+-------+----------+-------+-------+

| dept|countOfEmp|min\_sal|max\_sal|

+-------+----------+-------+-------+

| IT| 2| 2000| 5000|

| HR| 2| 3000| 4000|

|payroll| 2| 3500| 4500|

+-------+----------+-------+-------+